All et al Forty-year trends in the flux and concentration of phosphorus in British rivers

Aylin Civan[1], Fred Worrall[1*], Helen P. Jarvie[2], Nicholas J.K.Howden[3] and Tim P.Burt[4]

1. Dept. of Earth Sciences, Science Laboratories, South Road, Durham, DH1 3LE, UK.
2. Centre for Ecology and Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK.
3. Dept. of Civil Engineering, University of Bristol, Queens Building, Bristol, UK.
4. Dept. of Geography, Science Laboratories, South Road, Durham, DH1 3LE, UK.

Keywords: total reactive phosphorus; total phosphorus; change point analysis; sewage treatment.

Abstract

Given the importance of phosphorus (P) in the eutrophication of natural waters, this study considered the long-term time series of total phosphorus (TP) and total reactive phosphorus (TRP) in British rivers from 1974 to 2012. The approach included not only trend analysis of fluxes and concentrations but also change point analysis. TP and TRP concentrations and fluxes in British rivers have declined since the mid-1980s. Over the last decade of the record the majority of individual sites did show significant downward trends in TP and TRP concentrations but, in 28% of cases for TRP concentration and 14% of cases for TP concentration, the decadal trend was a significant increase. Out of 230 sites, 136 showed a significant step decrease in TRP concentration; no sites showed a significant step increase. The modal year for the step changes for both TRP concentration and flux was 1997. Step changes

[*] Corresponding author: Fred.Worrall@durham.ac.uk; tel. no. +44 (0)191 3342995; fax no. +44 (0)191 334 2301

are likely associated with improvements made at sewage treatment works to comply with the Urban Waste Water Treatment Directive (91/271/EEC). The decrease in TRP concentration due to the step change were in the range of 0.68% and 89% with a geometric mean of 22%, with the rest of the decrease accounted by long-term, persistent downward trend.

1. Introduction

Phosphorus (P) is an essential nutrient for the metabolic functioning of all forms of life, it plays a crucial role in controlling productivity in both terrestrial and aquatic ecosystems (Correll, 1998; Caraco, 2009). As an essential element for crop production phosphorus is vital for production of our food and bioresources; however inefficiencies in P use management and losses of P from agriculture and waste streams are a major source of impairments to water quality and security (Jarvie et al., 2015).

Phosphorus from point and diffuse sources are a major contributor to eutrophication and impairment of surface water quality at the global scale (Hecky & Kilham, 1988; Mainstone & Parr, 2002). Contributions of these sources differ depending on catchment characteristics such aspopulation and land use (Smith, et al., 1999). Point sources usually contain a high proportion of soluble and more biologically available phosphorus (Jarvie, et al., 2006) while diffuse sources are generally in particulate forms (P sorbed to soil particles) (EA, 2015). The contributions of agricultural diffuse P loadings (e.g. farmyard runoff, pig slurry, erosion from fields.) can be substantially higher than from urban sources such as sewage treatment works (STW) effluent (Comber et al, 2013; Naden et al, 2016), road runoff and septic effluent (Edwards & Withers, 2008). Equally, in some UK catchments STWs have been reported as the dominant source of P, eg. Iversen et al. (1997) and Parr and Mainstone (1997), but spatial budgets for P flux are generally lacking. Worrall et al. (2016) in their geographically weighted regression analysis of TRP and TP fluxes across the UK found significant roles for urban land use alongsisde and in comparison to significant roles for a range of soil types and agricultural land uses (Figure 1). For the flux of TP the most important source was urban land use which was found to be exporting 1.65 tonnes P/km^2 of urban land use (Worrall et al., 2016), but that study could not detail what were the specific sources were within urban land use.

A proportion of P loadings from diffuse and point sources within a catchment can accumulate in soils and aquatic sediments along transport pathways; this accumulated P can be remobilized or recycled, acting as a continuing source of P transport with residence times of years to decades (Sharpley, et al., 2014). This source of P has been referred to as "legacy-phosphorus" and can cause considerable delay in recovery of water quality impairment (Jarvie et al, 2013a; May et al., 2012; Powers, et al., 2016).

The European Commission's Urban Waste Water Treatment Directive (UWWTD - European Commission, 1991: 91/271/EEC) was brought in to restrict the pollution of natural waters by wastewater including limiting urban water as a source of P. As a result of the UWWTD numerous actions have been taken to reduce direct phosphorus inputs into rivers from sewage treatment works (STW), most notably tertiary treatment of effluent (Defra, 2002; Neal, et al., 2010a).

The UK lowlands have large population densities and thus rivers are highly vulnerable to P-sourced euthrophication (Neal, et al, 2010a). Mainstone & Parr (2002) have pointed out that catchments might be more sensitive to point sources such as STWs as they supply a similar flux of P even at low flows. Indeed, based upon observations from 54 monitoring sites, Jarvie et al. (2006) showed that even in rural areas with highly intensive agricultural P loadings, point P sources (wastewater) can represent a greater risk for river eutrophication linked to the dominant contributions of effluent P under baseflow conditions and periods of greatest ecological sensitivity – this should viewed in contrast to the work cited above (Comber et al, 2013; Naden et al, 2016). White and Hammond (2007) estimated that the total SRP (soluble reactive phosphorus) load of UK catchments was composed of 78% household, 13% agriculture, 4% industry and 6% background contributions; and for total TP load 73% household, 20% agriculture, 3% industry and 4% background contributions.

The UK has accomplished considerable progress on P remediation; however it still falls behind other EU members states such as Denmark, Sweden, Finland and the Netherlands where whole territories have been treated as sensitive areas, and also a large proportion of the countries like Germany and France have been designated the same way (Mainstone & Parr, 2002; IEEP, 1999). By 2002, only 2% of STWs in the UK had P-stripping installations (Foy, 2007). Muscutt and Withers (1996) carried out a study among 98 rivers in England and Wales and reported that 80% of the rivers were failing a target limit of of 0.1 mg/l mean orthophosphate concentration (DoE, 1993). However, the UK has increased investment and accelerated implementation of the UWWTD at STWs in the last decade (Bowes, et al., 2010): investment in England was almost doubled from £9600 M in the period 1990 - 2000 to £16100 M in the years between 2000 – 2015; with a total investment of £39126 M on STWs overall in the UK for the years between 1990 – 2015 (DEFRA, 2015). These actions have started to pay off with considerable reduction of phosphorus concentrations in many UK rivers (Kinniburgh & Barnett, 2010; Bowes, et al., 2009; Neal, et al., 2010c). Earl et al. (2014) found significant decreases in phosphate concentrations at the tidal limit in 68 out of 119 UK catchments that they considered between 1993 and 2003 which they ascribe to decreased use of polyphosphate in detergents and the introduction of the UWWTD. Nevertheless, the extent to which riverine phosphorus remediation will achieve the desired level of ecological improvement and also the timescales required for ecological recovery remains uncertain (Bowes, et al., 2010; Jarvie et al., 2013b).

The objective of this study was to understand the role of the UWWTD in reducing concentrations and fluxes of P at a national scale. The study will approach this objective by quantifying the concentration and flux over time of total TRP and TP at 230 monitoring sites across Britain.

2. Methodology

2.1. Study Sites

This study used datasets obtained from Harmonized Monitoring Scheme (HMS: Bellamy and Wilkinson, 2001). The HMS is a long-term river quality monitoring programme established in 1974 by the Department of Environment (DoE) and has been administrated by the Environment Agency since 1998. The programme comprises 270 river monitoring stations: data records of 230 HMS sites were suitable for this study. Of these 230 sites, 56 sites are located in Scotland and 174 sites are located in England and Wales (Figure 2). No data were available from Northern Ireland; therefore this study was restricted to Great Britain (GB) rather than the entire United Kingdom (UK). For inclusion in the monitoring programme, locations at the tidal limit of main rivers with an annual average discharge above 2 m^3/s or tributaries with an average annual discharge above 2 m^3/s were selected. With this discharge criterion, a good spatial coverage on the coast of England and Wales was achieved. However, in Scotland widespread coverage is not achieved because many of its west-coast rivers are too small for inclusion.

Four determinands within the database maintained by the HMS programme were used in this study: TRP concentration (mg P/l); TP concentration (mg P/l); instantaneous flow (m^3/s) and daily average flow (m^3/s). Due to the methodology used within the HMS monitoring programme (Simpson, 1980; DoE, 1972), the entries listed as orthophosphate concentration should be considered as TRP, since the methodology for orthophosphate measurement is based on colorimetric analysis of molybdate-reactive P on an *unfiltered* sample, and thus contains orthophosphate, and other easily-hydrolysable P fractions in both dissolved and particulate phases (Jarvie, et al, 2002, 2003). For the TP measurement, there is an additional acid-persulphate digestion step before the colorimetric analysis (DoE, 1972). The number of TP data available for analysis was much less than TRP data records - 40887 data points for TP compared to.118547 datapoints for TRP concentration.

In this study, both concentration and flux of TRP and TP were considered. Due to different monitoring agencies in charge, sampling frequencies and length of records varied between the sites. The length of records available for the study was from 1974 to 2012. Annual data were rejected at any site within any catchment where there were fewer than 12 samples in that year with the samples in separate months ($f<12$); in this way, a range of flow conditions would be sampled. Given this sampling frequency criterion, only 230 of the 270 sites within the HMS could be included in this study. Flux calculations were carried out using the method of Worrall et al. (2013). The method is based on the nature of the sources of variation within the flow and solute datasets and is a very simple method with a very low bias (8% for f = 1 per month) and a high accuracy (2% at f = 1 per month). The fluvial flux of a solute was estimated by the equation:

$$F = KE(C_i)Q_{total} \qquad \text{(i)}$$

where: Q_{total} = the total flow in a year (m^3/yr); $E(C_i)$ = the expected value of the sampled concentrations (mg/l); and K = constant for unit conversion (0.000001 for flux in tonnes). For the best results, the expected value of sampled concentration was based upon expected value of a gamma distribution. Flux calculations were made for both TRP and TP records at each HMS site, where the sampling frequency criterion was met and the total flow per year could be estimated from daily flow measurements – there were 4920 site-year combinations where TRP flux could be calculated during the study period while for TP flux there were just 2228 site-year combinations where the flux could be calculated..

2.2. Analysis of Variance and Covariance

Analysis of variance (ANOVA) was used to test the difference in TP and TRP fluxes between monitoring sites across time. Two factors were considered, henceforward referred to as *site* and as *year*. The site factor had 230 levels representing each monitored site and the year factor had 39 levels; one for each for the study period of 1974 – 2012. Analysis of variance was repeated using the water yield (annual average flow) as a covariate for each site.

ANOVA assumes that each population is normally distributed; therefore, normality of the datasets was checked prior to analysis using the Anderson-Darling test (Anderson & Darling, 1952). When any non-normality was found, the data (or covariate) were log-transformed before implementation of ANOVA: no further transformation was found to be required. Results are expressed as least square means (or marginal means) as they are the means controlled for all the factors and covariates. All results are reported at a significance level of $p < 0.05$ (95% probability of being different from zero). The results were expressed as least squares (or marginal) means within main effect plots; least square means are marginal means corrected for the influence of all other factors, interactions and covariates, to visualise the data.

The proportion of variance, or in other words the magnitude of the effect of each factor and covariate, was estimated by generalized omega square (ω^2) statistics (Olejnik & Algina, 2003). The ω^2 statistic is a different statistic than ANOVA's coefficient of determination (R^2) as that coefficient only explains the total variance in a model and does not give any information on individual contribution of factors to the variance.

2.3. Time Series and Trend Analysis

Trend analysis was used as a preliminary tool to obtain general descriptive information (positive, negative trends, or no trend) about each P time series for each catchment; therefore it was performed using linear trend analysis. The seasonal Kendall test (Hirsch et al., 1982) was used to assess the significance of any trend in the data sets and used to estimate the slope of any trend expressed as median annual change in the P concentration. The seasonal Kendall test is robust against departures from normality and resistant to outliers (Esterby, 1997). Only sites with at least 20 years of data between the study period of 1974-2012 were subjected to trend analysis for annual average concentration and flux data of TRP and TP. Also, a further trend analysis was performed on the last decade of the study period with the sites having 8 or more years of records between the years 2003-2012.

2.4. Change Point Analysis

In this study, preliminary visual inspection of the time series suggested that there were large step changes present in the series. To detect step changes in flux and concentration time series of TP and TRP, a non-parametric method Pettitt's test (Pettitt, 1979) was used. The Pettitt's test uses rank-based Mann-Whitney statistics $U_{t,N}$ comparing two independent sample sets $x_1, x_2, \ldots, x_t$ and $x_{t+1}, x_{t+2}, \ldots, x_N$ to test whether these sample sets are from the same population. The test statistic $U_{t,N}$ is calculated as follows:

$$U_{t,N} = U_{t-1,N} + \sum_{j=1}^{N} sgn(x_t - x_j) \qquad \text{for} \qquad t = 2, \ldots, N \qquad \text{(ii)}$$

and $$sgn(x_t - x_j) = \begin{cases} 1, & x_t > x_j \\ 0, & x_t = x_j \\ -1, & x_t < x_j \end{cases} \qquad \text{(iii)}$$

The step change is defined where $U_{t,N}$ has the maximum value, K_n:

$$K_n = Max|U_{t,N}| \qquad \text{(iv)}$$

The significance level of the step change is approximately:

$$P = exp\left(\frac{-6(K_n)^2}{n^3+n^2}\right) \qquad \text{(v)}$$

An enhanced probability estimation was suggested by Wilks (2006) for joint evaluation of repeated test results, as in the case of Pettitt's test, otherwise, familywise error (false detection rate – Ventura et al., 2004) will arise. Familywise error represents Type I errors (incorrect rejection of a true null hypothesis – false positive) in multiple hypothesis tests and, as more tests are performed, probability of Type I error increases.

To correct the familywise error, a new significance level is defined by using the method developed by Sidak (1967):

$$\alpha_{corrected} = 1 - (1 - \alpha)^{\frac{1}{N}} \qquad \text{(vi)}$$

where: α is the significance level or probability (α = 0.05 for Mann-Whitney U tests, 95% probability of a step change); $\alpha_{corrected}$ is the equivalent significance level that a test should be evaluated at; and N is the number of repeated tests.

For estimation of the effect size of any established step change, the Common Language Effect Size (CLES) method was used. In the CLES approach, the scores are ranked and all possible data pairs are compared for the compliance with the hypothesis, in the case of this study "the step change". As the name implies, the results are reported in a common language which is the percentage of pairs supporting the step change.

Correction of the familywise error has not been implemented in many studies that have used the Pettitt's test (e.g. Xu et al., 2014). Furthermore, with the Sidak correction, only the enhanced probability of Type I errors (false positives) are overcome; the probability of Type II errors (false negatives) should also be considered, and again, this has been lacking in many studies employing the Pettitt test (e.g. Zhang et al., 2014). For estimation of the probability of a false negative (β), statistical power analysis was performed.

Assuming effect sizes of 0.2, 0.5 and 0.8 with samples from 10 to 50 and assuming ratio of group sizes of 0.5, 0.66 and 0.75, *a priori* power analysis approach was conducted by comparing the asymptotic relative efficiency to a t-test based on Lehman's method using the G*Power 3.1 software (Faul *et al.*, 2007; http://gpower.hhu.de/). The acceptable power was set at 0.8 (a false negative probability β = 0.2). According to the power analysis, the probability of a false negative could be approximated as:

$$(1-\beta) = 0.008T + 0.057d + 0.51\frac{t}{T} - 0.45 \qquad r^2 - 0.899,\ n = 35 \qquad \text{(vii)}$$

$$\quad (0.002) \quad (0.06) \quad (0.14) \quad (0.08)$$

Where: T is number of years in the time series (up to 39 in this study); d is the effect size (0.0 to 1.0); and t is the larger number of years in the time series prior to or after the step change (a maximum of 19 years in this study). The values in brackets below the equation represent the standard errors in the coefficients and the constant term. Also, a significance level of 95% was taken into consideration for inclusion of the variables.

Equation (vii) shows that for the power analysis of annual records as considered here where the maximum value of T is 39, then for the statistical power to reach the acceptable threshold of 0.8 (80%), this would only occur for the largest T (longest time series) where the step change was in the middle of the record ($\frac{t}{T}$ = 0.5) and the effect size was large (d = 0.9). Therefore, it can be concluded that, although false positives can be eliminated from the Pettitt's test by use of the Sidak test, a high chance of false negatives will still remain.

Pettitt's test was applied to the annual average flux and concentration time series data of TRP and TP for the study period of 1974 - 2012.

3. Results

3.1. TRP Concentration

There were 118547 data points that could be paired for concentration and flow from 1974 to 2012 for TRP. The TRP concentration data had a median of 0.145 mg P/l with a 5^{th} to 95^{th} percentile range of 0.008 to 2.2 mg P/l. Prior to analysis, the Anderson-Darling test was performed and the data were found not to be normally distributed but log transformation was sufficient to normalise the data.

An ANOVA on TRP concentration records (Table 1) showed that both factors (site and year) were significant, with site being the most important factor both with and without the covariate. Inclusion of the log-transformed water yield as a covariate had only a negligible effect on explaining the variance on TRP concentration. *Post hoc* comparisons displayed significant differences between most of the sites. The magnitude of the variation between sites can be explained by the different soils, land use and point sources between catchments. Since the site factor had 230 levels the main effects would be very constrained, so data were investigated on a national level rather than on a river basin scale.

Main effects plots for TRP concentration with respect to the year factor (Figure 3) illustrates an overlapping structure for the two sets of concentration data with and without the covariate which indicates that inclusion of the covariate did not create a significant change in the analysis. The average TRP concentration has been declining since its peak in 1984; it has fallen from 0.16 mg/L in 1984 to 0.064 mg/L in 2012.

3.2. TRP flux

For the study period 1974-2012, the number of site-year combinations for which a TRP flux could be calculated was 4920 and the number of sites for which one year’s flux could be calculated varied from 16 in 1974 to 167 in 2011. Flux data were also checked for normality via Anderson-Darling test and found not to be normally distributed, and were therefore log-transformed before ANOVA. All factors were found to be significant at $p<0.05$ with site being the most important factor both with and without the covariate (Table 2). When the log-transformed water yield was included as the covariate, it significantly reduced the importance of the site factor. Also, the importance of the year factor was diminished by the inclusion of the covariate.

The main effects plot of the year factor with respect to TRP, both with and without the flow covariate, displays a fluctuating decrease over the course of the study (Figure 4). Inclusion of the covariate had a smoothing effect on the main effects of TRP flux by reducing the peak sizes and resulted in a clearer main effects profile for the flux. As for the TRP concentration (Figure 3), the TRP flux has been in decline since the mid-1980s confirming that the effect observed in Figure 4 is not due to hydroclimatic drivers such as changing river flows but does represent a real decline in the amount of phosphorus moving through the fluvial network. Worrall et al. (2016) have shown that the British national flux of TRP has been in decline since the mid-1980s but found a sharp decrease in 1993 and also a peak in 2002 that was not observed in the main effects plot in Figure 4.

3.3. TRP Trend Analysis

For sites with records over the entire study period 1974 to 2012, out of the 230 sites, there were 143 sites having a length of time series of 20 or more years for TRP concentration and flux. In the concentration time series, 116 sites had negative and 23 sites had positive trends while 4 sites had no significant trend. Among the 143 sites available for TRP concentration trend analysis, 84 of them had an annual average below 0.2 mg/l by the end of the study period while a further 76 sites had an annual average below 0.1 mg/l by the end of the study period. Similarly, for the flux TRP flux time series, 92 sites showed a negative trend while 41 sites showed a positive trend with 9 sites showing no significant trend. A map of the sites with negative TRP concentration trends (Figure 5a) illustrated that the sites with the largest declines were from the Midlands and South East regions of England whereas South West England, Wales and Scotland had very low rates of decline.

The largest significant decline in the TRP concentration was observed in time series of River Alt above Altmouth pumping station (National Grid Reference (NGR): SD2921105091) in which the TRP concentration and flux has been declining since 1979 and 1981, respectively and displaying a sharp peak in year 1990. The largest significant increase in TRP concentration time series was observed in the records of River Douglas at Wanes Blades Bridge (NGR: SD4758912612).

Trends in the Last Decade

There were 80 sites having 8 or more years of TRP concentration and flux data between years 2003-2012 – the last decade of the study period. For the concentration time series, 54 sites had significant negative trends, 22 sites had significant positive trends and the remaining 4 sites had no significant trend. By the end of the decade all the sites had an average TRP concentration below 0.2 mg/l and 79 sites had an average below 0.1 mg/l. The largest decline over the final decade of the study was the River Stour at Stourport Footbridge (NGR: SO8127070790). The site with the largest increase in TRP concentration time series in the last decade of the study period was the River Alyn at Ithels Bridge (NGR: SJ3902056230).

3.4. TRP Change Point Analysis

For the annual average concentration of TRP of the 230 sites where a record could be tested over the entire period of the study from 1974 to 2012, 136 sites for TRP showed a significant step change after familywise error correction; all the significant step changes were a step decline to a lower annual average. It is difficult to discern any particular pattern to the spatial distribution to the year when a significant step change occurred (Figure 5b) but most of the step changes occurred between 1993 and 1997; very few step changes were detected by 2003 – 2007. The modal year of the step change was 1997.

Common language effect sizes of the steps in TRP concentration ranged from 0.14 - 1.00 with a geometric mean of 0.89. In addition to CLES, actual sizes of the step changes were calculated by using the average concentrations for the year before and for the year after the step change. For TRP concentration, the actual sizes of the step changes were in the range 6.7 – 96% with a geometric mean of 49% (step changes were in the range between 0.001 mg/l and 2.36 mg/l with a geometric mean of 0.13 mg/l). Figure 5c show that step changes were rare for the rivers in Northern Scotland – an area of low population, extensive upland and livestock farming. The proportions of decrease in TRP concentration due to the step change were in the range of 0.68% and 89% with a geometric mean of 22%, with the rest of the decrease accounted by secular downward trend.

For the annual average TRP flux records, the number of sites having a significant step change was lower than for TRP concentration. After familywise correction, 74 sites for TRP flux were found to have step changes with common language effect sizes in the range 0.46 - 0.99 and a geometric mean of 0.87. The actual size of the step changes in flux were calculated to be in the range 5.8 – 97% with geometric mean of 50% for TRP (step changes were in the range between 0.33 ktonnes/yr and 1702 ktonnes/yr with a geometric mean of 38.9 ktonnes/yr). The majority of the step changes were in the period 1992– 1999; however, the modal year was 2000. Given that most of the concentration step changes also occurred in a similar period 1993 – 1997, it can be inferred that the real source of change is concentration and not river flow.

3.5. Total Phosphorus

TP Concentration

There were 40887 TP concentration data points that could be paired with flow records over the study period 1974 - 2012. The TP concentration had a median of 0.11 mg P/l with a 5^{th} to 95^{th} percentile range of 0.012 to 1.36 mg P/l. The Anderson-Darling test showed that the data distribution was not normal, and therefore the data were log-transformed – further transformation was not required.

The ANOVA results (Table 3) showed that both factors were significant with site factor being the most important factor for the TP concentration both with and without the covariate. Inclusion of the log-transformed water yield as the covariate increased the importance of the year factor and decreased the proportion of variance explained by the site factor.

Main effects plot of TP concentration with respect to the year factor with and without the covariate (Figure 6) displays an overlapping structure indicating that the inclusion of the covariate did not have a very large effect on ANOVA. The TP concentration peaked in 1985 and has been declining since that year; the least mean square TP concentration has decreased from 0.33 mg/L in 1985 to 0.10 mg/L in 2012. The main effects plot of TP concentration shows a sharp decrease in 1994 (Figure 6); it is possible that this is due to a change in sampling in that year. However, it should be pointed out that by using ANOVA including site and year factors, the time series of the least square means of the year factor over time should be independent of the site factor. Still the monitoring programme is not completely cross-classified with respect to sites and years of sampling and so the least squares means of the time factor could still reflect changes in sampling, equally differences between the years might be caused by specific events that occurred at particular sites in some years.

TP Flux

The number of site-year combinations for which a TP flux could be calculated was 2228 and the number of sites for which one year's TP flux could be calculated was between 1 in 1983 and 199 in 2012. The Anderson-Darling test suggested that TP Flux data were log-normally distributed and data were log-transformed prior to ANOVA.

Analysis of variance (Table 4) indicated that both factors were significant at $p<0.05$ with the site being the most important factor both with and without the covariate. When water yield was included in the analysis as the covariate, the importance of both site and year factors diminished.

The main effects plot of the year factor with respect to the TP flux both with and without the flow covariate (Figure 7) displays a fluctuating profile with flux decreasing since the mid-1980s. As with the main effect plot of TP concentration over time (Figure 6), there is a sharp decrease in the least squares mean for TP flux in 1994.

3.6. TP Trend Analysis

Overall Trends

For TP concentration and flux time series, there were only 8 sites having at least 20 years data for trend analysis. TP concentrations declined to below 0.2 mg/l in 6 of the sites and below 0.1 mg/l in 5 sites. All the sites showed significant negative trends in both concentration and flux except for one site: River Ness at Inverness (NGR: NH665445). The time series for the River Ness shows that both concentration and flux data of TP increased between the years 1994 - 2004 but a large spike in TP concentration in 2004 was of sufficient magnitude to distort the overall trend – we have no explanation for this spike.

Trends in the Last Decade

In the last decade of the study period, 95 sites had 8 or more years of TP concentration and flux data. For concentration time series, 13 sites showed significant positive trends, 78 sites showed significant negative trends and 3 sites did not show any significant trend. Among these sites, 93 of the sites declined to have an annual average below both 0.2 mg/l and 0.1 mg/l for TP concentration by the end of the study period. For the flux time series, 47 sites had significant positive trends, 44 had significant negative trends whereas 3 sites had no significant trend. The site with the steepest decline in TP concentration and flux time series was the River Stour at Stourport Footbridge (NGR: SO8127070790). The River Carnon at Devoran Bridge (NGR: SW7908739436) had the largest increase in TP concentration and flux time series in the last decade of the study period.

3.7. Change Point Analysis

For the annual average TP concentration, 31 sites were found to have significant step changes (after familywise correction) with all the step changes being to lower concentrations with common language effect sizes in the range of 0.80 - 1 with a geometric mean of 0.93 (Figure 8). Also, the actual sizes to the step changes were estimated to be in the range of 20 – 83% with geometric mean 50% for TP concentration (Magnitudes of changes were in the range between 0.013 mg/l and 1.72 mg/l with a geometric mean of 0.16 mg/l). Most of the step changes were in the period of 2003 – 2006 which is in contrast to the most frequent period of step changes found for the TRP concentration and flux; however, because of the scarcity of the TP records prior to 2002, most of the sites had very few data before the last decade of the study period (Figure 8).

For the annual average flux records, the number of sites having a significant step change is again lower than for TP concentration. After familywise error correction, only 4 sites had statistically significant step changes. The TP flux step changes had common language effect sizes in the range of 0.55 - 0.96 with a geometric mean of 0.79. The actual size of the step changes in flux were calculated to be in the range 32 - 81% with geometric mean of 50% for TP (size of changes were in the range between 9.2 tonnes/yr and 906 tonnes/yr with a geometric mean of 52.6 tonnes/yr).

For none of the time series considered was there a step change before 1982 (Figure 9). For TP concentration and flux there was no step change before 1996 and the modal years were 2005 and 2004 respectively.

3.8. Comparison to fertiliser inputs

Annual average concentration have declined since 1984 and 1985 for TRP and TP respectively (Figures 6 and 7). The British Survey of Fertilizer (Defra, 2015) shows that phosphate fertilizer usage in Great Britain peaked in year 1984 at 217530 tonnes P/yr, falling to 82387 tonnes/yr in 2012 which corresponds to a 62% decline (Figure 10). Despite the dramatic decline in fertilizer inputs, TRP flux did not decrease at the same rate. Over the same period, TRP flux declined from 33523 tonnes/yr in 1985 to 16481 tonnes/yr in 2012: a 50% decline. The ratio of TRP flux to fertilizer consumption was 0.15 in 1984 and 0.20 in 2012. On the other hand, TP flux has declined from 84288 ktonnes/yr in 1985 to 29643 ktonnes/yr in 2012, corresponding to a 64% decline; the ratio of TP flux to fertilizer consumption was 0.38 in 1984 and 0.36 in 2012. There are, of course, multiple sources of both TRP and TP to the GB river network but the changing balance between flux and fertiliser input illustrates that not only the proportion of sources could be changing, but also that accumulated or legacy P could be an increasingly important source (Waldrip et al., 2015). Surplus P has been noted in a range of environments, (e.g. global croplands - MacDonald et al., 2011) and that this has led to increased eutrophication of surface waters (e.g. Bennett et al., 2001). It has been proposed that this surplus can lead to prolonged P; leakage, which has been referred to as 'legacy' P (Sharpley et al., 2013; Jarvie et al 2013a). Haygarth et al. (2014) proposed that one approach to understanding whether P legacy would alter P concentrations and fluxes in surface waters would be to compile a full P budget for a catchment over time to explore trajectories of P accumulation and drawdown. Indeed, Powers et al. (2016) demonstrated for two large river basins (Thames, UK, Maumee, USA) that during the 1990s the net export from those catchments exceeded the inputs suggesting that today's river P fluxes reflect contributions from legacy stores of P.

There was a strong, and significant correlation apparent between UK P fertiliser use and average TRP concentration (Figure 10 - Annual average TRP conc. (mg P/l) = 0.00072*annual fertiliser input (ktonnesP/yr), n= 39, r^2 = 0.88, P= 0.00). There were also significant correlations between fertiliser inputs and annual average TRP flux, TP concentration and TP flux, but the correlation was strongest with the annual average TRP concentration. This significant correlation would imply that rivers leaving the UK would breach a standard of 0.1 mg P/l if the fertiliser input increased to over 139 ktonnes P/yr. However, such correlations will be misleading given the significant step changes observed for the individual time series (Figure 9). The secular trend in the national TRP and TP flux after 1985 could then be interpreted as representing the amalgamation of step changes in individual catchments.

4. Discussion

Understanding the driving forces behind these step changes is a complex issue; they cannot easily be attributed to linear drivers unless a threshold response can be justified. A step change could be caused by singular hydroclimatic events such as periods of drought or of sustained period of floods. On the other hand, droughts would be followed by a recovery and thus the effect sizes of the steps caused by droughts would be small. However, this study shows clear step changes with CLSE of approximately 89%. Also, several major droughts that affected the UK have been listed by Hannaford (2015) for the years 1990 - 1992, and 1995 - 1997 but none for 2000s, whereas there are many step changes encountered in the 2000s in the results presented here (Figure 9). Therefore, the step changes encountered in this study cannot readily be explained by droughts.

Comparing the geometric means of the effect sizes of concentration and flux records, it can be asserted that the actual step change is in concentration data and variations in flow can restrict step changes calculated for the flux records, resulting in fewer change points and lower effect sizes than the concentration records. The proportions of decrease in TRP concentration due to the step change were in the range of 0.68% and 89% with a geometric mean of 22%, with the rest of the decrease accounted by secular downward trend. Step changes can clearly be important contributors to overall decreasing trends therefore. Due to the lack of TP records in the monitoring database, the number of sites having significant step changes for TP concentration and flux are quite low and thus it is difficult to make interpretations on the step changes of TP.

Following the UWWTD, the Water Framework Directive (WFD) (Council of European Communities, 2000) was enacted to deliver improved water quality in a range of waterbodies. Phosphorus was indicated as as one of the main pollutants with an emphasis on eutrophication of sensitive areas defined as part of implementing the UWWTD. With respect to the UK, the Department of Environment, Food & Rural Affairs (DEFRA) reported that there were 588 sensitive areas in the UK comprising 19,466 km of rivers and canals and 2,737 km^2 of total catchment area (Defra, 2012). The targets for remediation of sensitive areas and elimination of euthrophication are to reduce annual average concentration of reactive P (RP: also referred to as orthophosphate or total reactive phosphorus

- Neal et al., 2010b) to between 0.02 mg P/l and 0.12 mg P/l; alkalinity is used to determine where in the acceptable range the SRP concentration should be (WFD UK TAG, 2008). Under the terms of UWWTD, sewage treatment works of <10000 population equivalent (p.e.) were required to install a P-stripping unit (as tertiary treatment) (Ferrier & Jenkins, 2009). STWs were also required to either meet the specified concentration limits for P in the final effluent (defined as 2 mg/l for 10000 to 100000 p.e. and 1 mg/l for over 100000 p.e.) or to remove 80% of the incoming P (Kinniburgh & Barnett, 2010).

Given that most of the step changes in TRP concentration were encountered in the period of 1993 - 1997 (Figure 9) and assuming that the 1995-1997 drought had limited impact, these step changes can either be explained by a significant decrease in fertilizer inputs since 1993, or as a result of implementation of the UWWTD. Spatial distribution maps indicated that TRP concentration step changes encountered in the period 1993-1997 were mostly from urbanized regions such as the Midlands, North West and South East England, whereas in less urbanized regions like Northern Scotland, Wales and South West England mostly have step changes in different periods or has seen no change. Therefore, it can be proposed that these step changes were brought about by the implementation of the UWWTD. This is in agreement with other studies from individual British rivers which have assessed changes in P concentrations, fluxes and eutrophication risk since the 1990s (e.g., Bowes et al 2010, 2011, Whitehead et al, 2013; Earl et al, 2014; Bussi et a, 2016, Tappin et al., 2016). Worrall et al. (2016) could show that urban land use was the important contributor to the explanation of TP flux across the UK but did not consider change in this urban source over time. This study, however, for the first time demonstrates the integrated effects of these mitigation measures on riverine P fluxes at the national scale over time. The UWWTD required tertiary treatment to strip P from treatment works greater than 15000 population equivalent (p.e.) by the 31 December 2000. Indeed, this study can show that after the implementation of the UWWTD in 1992, the modal year for change was in 1997, i.e. before deadline of 2000. At that date the UK was 90% compliant with the requirement: by the end of 2007 it was 99.9% compliant (Defra, 2012). For treatment works between 2000 and 15000 p.e. the Directive required provision of secondary treatment by end of 2005. By the end of 2005, the UK was over 99% compliant, and indeed, this study has shown that within the decade 2003 to 2012, 24 out of the 28 significant changes occurred prior to 2005, but only four after 2005. Finally, designated sensitive areas require tertiary treatment (e.g. phosphorus stripping). The UK currently has 588 sensitive areas totalling 19,466 km of river channel with a total catchment area of 2737 km^2. It should be noted that although this study has associated step changes with changes at STWs we cannot comment on, or include in the analysis, the relative, or absolute, locations of the monitoring sites from the STWs in each catchment. Therefore, the scale of the step change demonstrated at each site that might be controlled by the distance from the nearest STW that has undergone a change in treatment technique.

Worrall et al. (2014) showed that changes in POM were being driven by improvements in waste water treatment after the implementation of the UWWTD (European Commission, 1991). Numerous actions have been taken to reduce direct phosphorus inputs into rivers from STWs (Defra, 2002; Neal, et al., 2010a), including provision of secondary treatment (eg. activated sludge process) and installation of phosphorus stripping units as tertiary treatment. Nevertheless, it should be noted that concentrations of TRP and TP started decrease in the mid-1980s, before the implementation of Urban Waste Water Treatment Directive.

In England since 1953 there have been 673 discharge consents made upon sewage treatment works that concerned the discharge of phosphorus from the works to surface waters, covering 187 separate locations. In all cases the new discharge consent asked for a lower phosphorus concentration although they varied as to whether the discharge consents concerned the maximum or the mean value of the phosphorus concentration in the discharge. The timing of the new discharge consents shows that the modal year of the consent being issued was 1989 (before the UWWTD came into force); the second most important year was 1997. These discharge consents may therefore have contributed to pre-UWWTD declines as well as the impact of declining fertilizer use.

5. Conclusions

Mean total phosphorus (TP) concentration for river waters leaving the UK has declined by 69% since a peak in 1985 – 60% decline for mean total reactive phosphorus (TRP) concentration.

In the decade (2003 to 2012) there was a significant decrease trend in TP concentration at 82% of the sites considered and a significant increases were only observed at 14% of 95 sites. Significant step changes were observed in many of the catchment records. Since 1974, of 230 sites with sufficient TRP concentration records, 136 showed a significant step decrease with a modal year of 1997, a mean common language effect size of 89% and an actual effect size of a 49% decrease. The step changes can be ascribed to actions taken at sewage treatment as part of the Urban Wastewater treatment directive (UWWTD). However, not all the decrease observed in TP or TRP can be ascribed to step changes or actions at sewage treatment works. The proportions of decrease in TRP concentration due to the step change were in the range of 0.68% and 89% with a geometric mean of 22%, with the rest of the decrease accounted for by long-term, persistent change.

References

Anderson, T. W., & Darling, D. A. (1952). Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic. *The Annals of Mathematical Statistics, 23*(2), 193-212.

Bellamy, D., & Wilkinson, P. (2001). OSPAR 98/3: an environmental turning point or a flawed decision? *Marine Pollution Bulletin, 42(2)*, 87-90.

Bennett, E.M., S.R.Carpenter and N.F.Caraco (2001) Human impact on erodible phosphorus and eutrophication: a global perspective. Bioscience, 51, 227-234.

Bowes, M. J., Neal, C., Jarvie, H. P., Smith, J. T., & Davies, H. N. (2010). Predicting phosphorus concentrations in British rivers resulting from the introduction of improved phosphorus removal from sewage effluent. *Science of the Total Environment, 408,* 4239–4250.

Bowes, M. J., Smith, J., Jarvie, H. P., Neal, C., & Barden, R. (2009). Changes in point and diffuse source phosphorus inputs to the River Frome (Dorset, UK) from 1966 to 2006. *Science of the Total Environment, 407*(6), 1954-1966.

Bowes, M.J., Smith, J.T., Neal, C., Leach, D.V., Scarlett, P.M., Wickham, H.D., Harman, S.A., Armstrong, L.K., Davy-Bowker, J., Haft, M., Davies, C.E., 2011. Changes in water quality of the River Frome (UK) from 1965 to 2009: Is phosphorus mitigation finally working? Science of the Total Environment 409, 3418-3430.

Bussi, G., Whitehead, P.G., Bowes, M.J., Read, D.S., Prudhomme, C., Dadson, S.J., 2016. Impacts of climate change, land-use change and phosphorus reduction on phytoplankton in the River Thames (UK). Science of the Total Environment 572, 1507-1519.

Caraco, N. (2009). Phosphorus. In G. E. Likens, *Encyclopedia of Inland Waters* (pp. 73-78). Elsevier.

Comber, S., Gardner, M., Georges, K., Blackwood, D., Gilmour, D., 2013. Domestic source of phosphorus to sewage treatment works. Environmental Technology 34, 1349-1358.

Correll, D. L. (1998). The Role of Phosphorus in the Eutrophication of Receiving Waters: A Review. *Journal of Environmental Quality, 27*, 261-266.

Council of European Communities. (2000). Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. *L 327*, P. 0001-0073.

Defra. (2002). *The Government's Strategic Review of Diffuse Water Pollution from Agriculture in England: Agriculture and Water: A Diffuse Pollution Review.* London: Department for Environment, Food and Rural Affairs.

Defra. (2012). *Waste water treatment in the United Kingdom – 2012* . London: Department for Environment, Food and Rural Affairs .

Defra (2015). *The British Survey of Fertilizer Practise: Fertilizer Use on Farm Crops for Crop Year 2014.* Crown Copyright 2014.

DoE: Department of the Environment. (1972). Analysis of Raw, Potable and Wastewaters. *HMSO*.

DoE: Department of the Environment. (1993). *Methodology for identifying sensitive areas (Urban Wastewater Directive) and designating vulnerable zones (Nitrates Directive) in England and Wales.* Department of the Environment, Ministry of Agriculture Fisheries and Food, Welsh Office.

Earl, T.J., Upton, G.J.G., Nedwell, D.B., 2014. UK catchment nutrient loads 1993-2003, a new approach using harmonised monitoring scheme data: temporal changes, geographical distribution, limiting nutrients and loads to coastal waters. Environmental Science-Processes & Impacts 16, 7, 1646-1658.

Edwards, A., & Withers, P. (2008). Transport and delivery of suspended solids, nitrogen and phosphorus from various sources to freshwaters in the UK. *Journal of Hydrology, 350,* 144 -153.

EFMA. (2000). *Understanding phosphorus and its use in agriculture – phosphorus.* Brussels: European Fertiliser Manufacturers' Association (EFMA).

European Commission. (1991). Urban Waste Water Treatment Directive. Council Directive 91/271/EEC.

Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, **39,** 175-191.

Ferrier, R., & Jenkins, A. (Eds.). (2009). *Handbook of Catchment Management.* Wiley-Blackwell.FWR: Foundation of Water Research. (2005). Information Note FWR-WFD16. *Sources of Pollution - Overview*.

Foy, R. H. (2007). Variation in the reactive phosphorus concentrations in rivers of northwest Europe with respect to their potential to cause eutrophication. *Soil Use Manage, 23,* 195–204.

Hannaford, J. (2015). Climate-driven changes in UK river flows: A review of the evidence. *Progress in Physical Geography, 39*(I), 29–48.

Haygarth, P.M., H.P.Jarvie, S.M.Powers, A.N.Sharpley, J.J.Elser, J.Shen, H.M.Peterson, N-I, Chan, N.J.K. Howden, T.P.Burt, F.Worrall, F.Zhang and X.Liu (2014) Sustainable phosphorus management and the need for a long-term perspective: the legacy hypothesis. *Environmental Science & Technology,* 48, 8417-8419.

Hecky, R. E., & Kilham, P. (1988). Nutrient limitation of phytoplankton in freshwater and marine environments: A review of recent evidence on the effects of enrichment. *Limnology and Oceanography, 33*(4), 796–822.

Institute of European Environmental Policy (IEEP). (1999). *The implementation of the 1991 EU Urban Waste Water Treatment Directive and its role in reducing phosphate discharges - summary of report.* SCOPE Newsletter No. 34.

Iversen, T.M., Kjeldsen, K., Kristensen, P., de Haan, B., van Oirschot, M., Parr, W., & Lack, T. (1997). Integrated environmental assessment on eutrophication. A pilot study. NERI Technical Report 307, Natural Environmental Research Institue.

Jarvie, H. P., Neal, C., & Withers, P. J. (2006). Sewage-effluent phosphorus: A greater risk to river eutrophication than agricultural phosphorus? *Science of the Total Environment, 360*, 246–253.

Jarvie, H.P., Neal, C., Withers, P.J.A., Robinson, A., Salter, N., 2003. Nutrient water quality of the Wye catchment, UK: exploring patterns and fluxes using the Environment Agency data archives. Hydrology and Earth System Sciences 7, 722-743.

Jarvie, H.P., Sharpley, A.N., Spears, B., Buda, A.R., May, L., Kleinman, P.J.A., 2013a. Water Quality Remediation Faces Unprecedented Challenges from "Legacy Phosphorus". Environmental Science & Technology 47, 8997-8998.

Jarvie, H.P., Sharpley, A.N., Withers, P.J.A., Scott, J.T., Haggard, B.E., Neal, C., 2013b. Phosphorus Mitigation to Control River Eutrophication: Murky Waters, Inconvenient Truths, and "Postnormal" Science. Journal of Environmental Quality 42, 295-304.

Jarvie, H.P., Sharpley, A.N., Flaten, D., Kleinman, P.J.A., Jenkins, A., Simmons, T., 2015. The Pivotal Role of Phosphorus in a Resilient Water-Energy-Food Security Nexus. Journal of Environmental Quality 44, 1049-1062.

Jarvie, H. P., Withers, P. J., & Neal, C. (2002). Review of robust measurement of phosphorus in river water: sampling, storage, fractionation and sensitivity. *Hydrology and Earth System Sciences*, *6*, 113-132.

Kinniburgh , J. H., & Barnett, M. (2010). Orthophosphate concentrations in the River Tham es: reductionsin the past decade. *Water and Environment Journal, 24,* 107–115.

MacDonald, G.K., E.M.Bennett, P.A.Potter and N.Ramankutty (2008) Agronomic phosphorus imbalances across the world's croplands. Proceedings of the National Academy of Sciences, 108, 7, 3086-3091.

Mainstone, C. P., & Parr, W. (2002). Phosphorus in rivers - ecology and management. *The Science of the Total Environment, 282-283*, 25-47.

May, L., Defew, L.H., Bennion, H. and Kirika, A. (2012). Historical changes (1905-2005) in external phosphorus loads to Loch Leven, Scotland, UK. *Hydrobiologia*, 681 (1). 11-21. 10.1007/s10750-011-0922-y

Muscutt, A. D., & Withers, P. J. (1996). The phosphorus content of rivers in England and Wales. *Water Research*, *30*(5), 1258-1268.

Naden, P., Bell, V., Carnell, E., Tomlinson, S., Dragosits, U., Chaplow, J., May, L., Tipping, E., 2016. Nutrient fluxes from domestic wastewater: A national-scale historical perspective for the UK 1800-2010. Science of the Total Environment 572, 1471-1484.

Neal, C., Jarvie, H. P., Wade, A. J., Neal, M., Wyatt, R., Wickham, H., . . . Hewitt, N. (2004). The water quality of the LOCAR Pang and Lambourn catchments. *Hydrology and Earth Sytem Sciences*, *8*(4), 614-635.

Neal, C., Jarvie, H. P., Withers, P. J., Whitton, B. A., & Neal, M. (2010a). The strategic significance of wastewater sources to pollutant phosphorus levels in English rivers and to environmental management for rural, agricultural and urban catchments. *Science of the Total Environment*, *408*, 1485-1500.

Neal, C., Martin, E., Neal, M., Hallett, J., Wickham, H. D., Harman, S., Keay, D. (2010c). Sewage effluent clean-up reduces phosphorus but not phytoplankton in lowland chalk stream (River Kennet, UK) impacted by water mixing from adjacent canal. *Science of The Total Environment*, *408*(22), 5306–5316.

Olejnik, S., & Algina, J. (2003). Generalized Eta and Omega Squared Statistics: Measures of Effect Size for Some Common Research Designs. *Psychological Methods*, *8*(4), 434-447.

Parr, W., & Mainstone, C.P. (1997). The Pevensey Levels data to support its inclusion as a candidate eutrophic sensitive area. Environment Agency Southern Region.

Pettitt, A. (1979). A non-parametric approach to change point analysis. *Applied Statistics, 28*, 126-135.

Powers, S., Bruulsema, T., Burt, T., Chan, N., Elser, J., Haygarth, P., . . . Zhang, F. (2016). Long-term accumulation and transport of anthropogenic phosphorus in three river basins. *Nature Geoscience, 9*, 353–356.

Sharpley, A., Jarvie, H., Buda, A., May, L., Spears, B., & Kleinman, P. (2013). Phosphorus Legacy: Overcoming the Effects of Past Management Practices to Mitigate Future Water Quality Impairment. *Journal of Environmental Quality, 42*, 1308–1326.

Sidak, Z. K. (1967). Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. *Journal of the American Statistical Association, 62*(318), 626–633.

Simpson, E. A. (1980). The Harmonization of the monitoring of the quality of rivers in th United Kingdom. *Hydrological Science Bulletin, 25*(1), 13-23.

Smith, V. H., Tilman, G. D., & Nekola, J. C. (1999). Eutrophication: impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. *Environmental Pollution, 100*, 179-196.

Tappin, A.D., Comber, S., Worsfold, P.J., 2016. Orthophosphate-P in the nutrient impacted River Taw and its catchment (SW England) between 1990 and 2013. Environmental Science-Processes & Impacts 18, 690-705.

Ventura, V., Paciorek, C., & Risbery, J. (2004). Controlling the proportion of falsely rejected hypotheses when conducting multiple tests with climatological data. *Journal of Climate, 17*, 4343–4356.

WFD UK TAG. (2008). *UK Environmental Standards and Conditions (Phase 1).* UK Technical Advisory Group on the Framework Directive. Retrieved from www.wfduk.org

White, P. J., & Hammond, J. P. (2007). The Sources of Phosphorus in the Waters of Great Britain. *Journal of Environmental Quality, 38*(1), 13-26.

Whitehead, P.G., Crossman, J., Balana, B.B., Futter, M.N., Comber, S., Jin, L., Skuras, D., Wade, A.J., Bowes, M.J., Read, D.S., 2013. A cost-effectiveness analysis of water security and water quality: impacts of climate and land-use change on the River Thames system. Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences 371

Wilks, D. (2006). On "field significance" and the false detection rate. *Journal of Applied Meteorology and Climatology, 45*, 1181–1189.

Worrall, F., Burt, T. P., & Howden, N. J. (2013). Assessment of sample frequency bias and precision in fluvial flux calculations – an improved low bias estimation method. *Journal of Hydrology, 503*, 101-110.

Worrall, F., T.P.Burt, and N.J.K.Howden (2014). The fluvial flux of particulate organic matter from the UK: quantifying in-stream losses and carbon sinks. *Journal of Hydrology*, 519, 611-625.

Worrall, F., Jarvie, H.P., Howden, N.J.K., Burt, T.P. (2016). The fluvial flux of total reactive and total phosphorus from the UK in the context of a national phosphorus budget: comparing UK river fluxes with phosphorus trade imports and exports. *Biogeochemistry*130, 1-2, 31-51.

Xu, K., Lian, J., & Bin, L. (2014). Joint Probability Analysis of Extreme Precipitation and Storm Tide in a Coastal City under Changing Environment. *Journal Plos One, 9*(10), e109341.

Zhang, Q., Gu, X., Singh, V. P., Xiao, M., & Xu, C. (2014). Stationarity of annual flood peaks during 1951–2010 in the Pearl River basin, China. *Journal of Hydrology, 519*, 3263–3274.

Figure 1. Comparison of: a) land use (given as percentage arable land use per km^2); and b) TP flux export per km^2 for Great Britain from Worrall et al. (2016).

Figure 2. . Location of HMS monitoring sites used in the study.

Figure 3. Main effects plot for the annual average TRP concentration with and without the covariate over the study period 1974 – 2012.

Figure 4. Main effects plot for the annual average TRP flux with and without the covariate over the study period 1974 – 2012.

Figure 5. Spatial distribution map of: a) TRP concentration trends for the overall study period; b) the year before any significant step change; and c) the common language effect size (CLSE) of the significant step changes.

Figure 6. Main effects plot for the annual average TP concentration with and without the covariate over the study period 1974 – 2012.

Figure 7. Main effects plot for the annual average TP flux with and without the covariate over the study period 1974 – 2012.

Figure 8. Spatial distribution map of: a) the year before any significant step change; and b) the common language effect size (CLSE) of the significant step changes.

Figure 9. Comparisons between step change years in TRP and TP concentration and flux time series records.

Figure 10. Overall phosphate fertiliser application for the UK (ktonnes P - Defra, 2015) in comparison to the annual least squares means of the TRP concentration.

Table 1. Results of ANOVA and ANCOVA on Total Reactive Phosphorus (TRP) Concentration.

Factor or Covariate	Without Covariate		With Covariate	
	P-value	Proportion of variance (ω^2)	P-value	Proportion of variance (ω^2)
ln(Water Yield)	-	-	0.003	0.015
Site	0	96.57	0	96.52
Year	0	2.93	0	2.95
Error	-	0.005	-	0.52

Table 2. Results of ANOVA and ANCOVA on Total Reactive Phosphorus (TRP) Flux.

Factor or Covariate	Without Covariate		With Covariate	
	P-value	Proportion of variance (ω^2)	P-value	Proportion of variance (ω^2)
ln(Water Yield)	-	-	0	9.25
Site	0	96.94	0	88.25
Year	0	2.12	0	1.92
Error	-	0.009	-	0.58

Table 3. Results of ANOVA and ANCOVA on Total Phosphorus (TP) Concentration.

Factor or Covariate	Without Covariate		With Covariate	
	P-value	Proportion of variance (ω^2)	P-value	Proportion of variance (ω^2)
ln(Water Yield)	-	-	0.135	0.006
Site	0	95.38	0	95.21
Year	0	3.35	0	3.43
Error	-	1.27	-	1.35

Table 4. Results of ANOVA and ANCOVA on Total Phosphorus (TP) Flux.

Factor or Covariate	Without Covariate		With Covariate	
	P-value	Proportion of variance (ω^2)	P-value	Proportion of variance (ω^2)
ln(Water Yield)	-	-	0	16.16
Site	0	94.01	0	79.44
Year	0	2.94	0	2.70
Error	-	3.04	-	1.69